Mismatch repair and cancer

SIR — Last March, a *Science* editorial hailed the discovery of a 'new cancer gene', linked to the hereditary nonpolyposis colorectal carcinoma (HNPCC) and localized to the human chromosome 2p16-15. The phenotype of the HNPCC cells appeared to be the instability of microsatellite DNA and other short repeats, and the accumulation of a large number of mutations scattered over the entire genome[1-3]. This looked like a typical hallmark of a mutator gene at work and it did not take long before a link was made between the colorectal carcinoma phenotype and that of *Saccharomyces cerevisiae* mutants, *pms1*, *mlh1* and *msh2*, which are all deficient in mismatch correction[4].

Further light was cast on the subject by Kolodner and colleagues[5], who described the localization of the human *msh2* homologue, *hMSH2*, to chromosome 2p22-21 and demonstrated that this gene does indeed encode a mutator function, as overexpression of *hMSH2* in *Escherichia coli* gave rise to 10-fold higher mutation rates. Two weeks later, Leach and 36 coauthors[6] followed up with a convincing study, reporting the localization of the *hMSH2* gene to chromosome 2p16. These authors also identified mutations, both germline and somatic, in the DNA of HNPCC patients. Both sets of authors concluded that *hMSH2* is the HNPCC gene and suggested that the protein encoded by *hMSH2* is probably involved in DNA mismatch repair. Modrich and colleagues[7] substantiated these claims by showing that HNPCC-derived cell line H6 cannot repair mismatches *in vitro* nor *in vivo*. We can provide independent confirmation that *hMSH2* encodes a mismatch-binding protein.

In 1988, we reported the identification of a human (HeLa) cell protein, GTBP, which could bind with high affinity to oligonucleotide duplexes containing G/T mispairs[8]. This protein which also recognizes mispairs other than G/T when purified[9] appeared to possess no enzymatic 'repair' activity on mismatch-containing DNA duplexes *in vitro*. But GTBP does have ATPase and helicase activities, which is consistent with the properties of MutS, a mismatch-binding protein of *E. coli*[10]. This was a very interesting finding, as Crouse and coworkers[11] reported the identification of a gene, *Rep-3*, divergent from the mouse *Dhfr* locus, which contained an open reading frame very similar to the bacterial *mutS* and which also appeared to have a human counterpart, *DUG* (ref. 12). This serendipitous discovery rang numerous bells in the world of mismatch repair and, thanks to PCR, led to the rapid identification of a family of *mutS*-homologous, or *msh*, genes in yeast, mouse and human genomes (see ref. 13 and references therein).

Proteolytic digestion of the purified 100,000 GTBP polypeptide with trypsin, followed by sequencing of the generated fragments, yielded ten peptides: FR27-ECVLPGGETAGDMGK; FR28-LTSLNEEYTK; FR30-GGILITER; FR39-MNFESFVK; FR40-IIQEFLSK; FR43-ENDWYLAYK; FR45-GDFYTAHGEDALLAA; FR48-LLSAQFGYYF; FR52-ETLQLESAAEVGF; and FR60-ALELEEFQYIGESQGY. Using degenerate oligonucleotides based on the sequence derived from peptide FR60 to probe a lambda-ZAP HeLa complementary DNA library (Stratagene), we isolated a partial cDNA clone which contains approximately 1,000 nucleotides of the C-terminal part of the GTBP open reading frame. The DNA sequence of this partial clone matches that of *hMSH2* (refs 5, 6). In addition, a consensus ATPase site and all the ten peptides above could be located in the translated *hMSH2* sequence, as shown in the figure. We therefore conclude that, based on the evidence presented above, *hMSH2* encodes GTBP.

As already stated, this latter protein appears to be a functional homologue of the *E. coli* mismatch-binding protein MutS, which plays a key role in the initiation of methyl-directed mismatch repair in this organism. Given that the principal roles of MutS are those of the recognition and binding of base–base mismatches[10] and small loops[14] as well as that of recruitment into the 'repairosome' of the other participating proteins MutL, MutH and others, and given that the absence or malfunction of the MutS protein reduces the fidelity of replication of this organism by three orders of magnitude[15], it is easy to understand the mutator phenotype of HNPCC cells where the *hMSH2* gene is non-functional.

The *hMSH2* story is clearly just the tip of an iceberg. One need only look at the complexity of the *E. coli* MutHLS methyl-directed mismatch repair system[16] to realize that a phenotype similar to HNPCC could be due to any one of a number of mutations in the large number of genes involved in mismatch correction. All that remains now is to identify them.

Fabio Palombo
Melya Hughes
Josef Jiricny
Istituto di Richerche di Biologia Molecolare, P. Angeletti, Via Pontina Km 30,600, 00040 Pomezia, Italy
Oanh Truong
Justin Hsuan
Ludwig Institute for Cancer Research, University College London Medical School, Courtauld Building, 91 Riding House Street, London W1P 8BT, UK

MAVQPKETLQLESAAEVGFVRFFQGMPEKPTTTVRLFDRGDFYTAHGEDALLAAREVFKTQGVIKYMGPAG
AKNLQSVVLSKMNFESFVKDLLLVRQYRVEVYKNRAGNKASKENDWYLAYKASPGNLSQFEDILFGNND
MSASIGVVGVKMSAVDGQRQVGVGYVDSIQRKLGLCEFPDNDQFSNLEALLIQIGPKECVLPGGETAGDMG
KLRQIIQRGGILITERKKADFSTKDIYQDLNRLLKGKKGEQMNSAVLPEMENQVAVSSLSAVIKFLELLSDDS
NFGQFELTTFDFSQYMKLDIAAVRALNLFQGSVVDTTGSQSLAALLNKCKTLKDKDLLTSGLSSLSWIRTRIE
ERLNLVEAFVEDAELRQTLQEDLLRRFPDLNRLAKKFQRQAANLQDCYRLYQGINQLPNVIQALEKHEGKHQ
KLLLAVFVTPLTDLRSDFSKFQEMIETTLDMDQVENHEFLVKPSFDPNLSELREIMNDLEKKMQSTLISAARD
LGLDPGKQIKLDSSAQFGYYFRVTCKEEKVLRNNKNFSTVDIQKNGVKFTNSKLTSLNEEYTKNKTEYEEA
QDAIVKEIVNISSGYVEPMQTLNDVLAQLDAVVSFAHVSNGAPVPYVRPAILEKGQGRIILKASRHACVEVQD
EIAFIPNDVYFEKDKQMFHIITGPNMGGKSTYIRQTGVIVLMAQIGCFVPCESAEVSIVDCILARVGAGDSQLK
GVSTFMAEMLETASILRSATKDSLIIIDELGRGTSTYDGFGLAWAISEYIATKIGAFCMFATHFHELTALANQIP
TVNNLHVTALTTEETLTMLYQVKKGVCDQSFGIHVAELANFPKHVIECAKQKALELEEFQYIGESQGYDIME
PAAKKCYLEREQGEKIIQEFLSKVKQMPFTEMSEENITIKLKQLKAEVIAKNNSFVNEIISRIKVTT

Amino-acid sequence of hMSH2, reproduced from ref. 6. GTBP peptides are boxed. Only a single deviation from the published sequence was found, an Asp to Leu mutation in peptide FR-48. The ATPase consensus sequence is underlined. (Note that the postulated hMSH2 open reading frame published in reference 5 is missing the N-terminal 25 amino acid residues. This is due to a difference in the cDNA sequence, whereby the true ATG initiation codon appears as ATC and methionine 26 was thus taken as the translation start site).

1. Aaltonen, L. A. *et al. Science* **260**, 812–818 (1993).
2. Thibodeau, S. N., Bren, G. & Schaid, D. *Science* **260**, 818–821 (1993).
3. Ionov, Y., Peinado, M. A., Malkhosyan, S., Shibata, D. & Perucho, M. *Nature* **363**, 558–561 (1993).
4. Strand, M., Prolla, T. A., Liskay, R. M. & Petes, T. D. *Nature* **365**, 274–276 (1993).
5. Fishel, R. *et al. Cell* **75**, 1027–1036 (1993).
6. Leach, F. S. *et al. Cell* **75**, 1215–1225 (1993).
7. Parsons, R. *et al. Cell* **75**, 1227–1236 (1993).
8. Jiricny, J., Huges, M., Corman, N. & Rudkin, B. B. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8860–8864 (1988).
9. Hughes, M. & Jiricny, J. *biol. Chem.* **267**, 23876–23882 (1992).
10. Su, S.-S. & Modrich, P. *Proc. natn. Acad. Sci. U.S.A.* **83**, 5057–5061 (1986).
11. Linton, J. P. *et al. Molec. cell. Biol.* **9**, 3058–3072 (1989).
12. Fuji, H. & Shimada, T. *J. biol. Chem.* **264**, 10057–10064 (1989).
13. New, L., Liu, K. & Crouse, G. F. *Molec. gen. Genet.* **239**, 97–108 (1993).
14. Parker, B. O. & Marinus, M. G. *Proc. natn. Acad. Sci. U.S.A.* **89**, 1730–1734 (1992).
15. Modrich, P. *A. Rev. Biochem.* **56**, 435–466 (1987).
16. Modrich, P. *A. Rev. Genet.* **25**, 229–253 (1991).

© 1994 Nature Publishing Group